Economics Bulletin

Volume 30, Issue 4

# Social networks and employment in India

Tushar K. Nandi
*ADIS, Université Paris-Sud*

## Abstract

We investigate the influence of social networks on employment. Using data from India, we estimate the effect of caste based social networks on employment. We use a methodology that allows us to control for several omitted variable biases that often confound network effect. Our results indicate that caste based social networks are important determinant of employment in India. The implication of our findings is that a policy of positive discrimination in labour market for disadvantaged caste is able to generate additional benefit through network channels.

**Citation:** Tushar K. Nandi, (2010) "Social networks and employment in India", *Economics Bulletin*, Vol. 30 no.4 pp. 2769-2778.
**Submitted:** Jun 18 2010. **Published:** October 26, 2010.

## 1. Introduction

Over the past decade economists have increasingly recognised the importance of social networks for economic and social outcomes. Social networks affect individual behaviours through two channels – information and norms (Bertrand et al. 2000). The information channel works by affecting an individual's knowledge through others' behaviour. The norm channel emphasises how an individual's preferences can be affected by the behaviour of others, either directly by affecting taste or indirectly via social pressure. Several empirical studies have focused on how social networks affect a variety of outcomes – job search (Montgomery 1991), education (Coleman et al. 1966), unemployment (Akerlof 1980), employment (Burns et al. 2009) to list only a few. In this paper we study the effect of caste based social networks on employment in India.

In India, caste is highly correlated with socio-economic status. This correlation implies that the effect of caste on an outcome is difficult to separate from the effect of socio-economic status of the caste. This difficulty is particularly apparent in the estimation of caste based network effects as individuals with different socio-economic status are likely to have different levels of reach in any network. Upper caste have better economic outcomes than non-upper castes (Scheduled caste, Scheduled tribe and Other backward caste). Non-upper castes constitute more than 50% of Indian population. Their economic backwardness has been subject of policy and political debate throughout the Indian history. Economic disadvantage of non-upper caste has often been attributed to discrimination leading to occupational segregation and denial to access to education. In order to equalise economic opportunities between castes, the constitution of India mandates reservation of places in higher education institutes, jobs in public sectors, and seats in regional and national legislatures for Scheduled caste, Scheduled tribe and Other backward caste. This reservation policy has been in place since its inception after India's independence in 1947 in order to fight the persistent disadvantage of non-upper castes. In this paper, we explore the possibility that caste based social networks can be a factor determining the persistence of inequality. In particular, we study whether caste based social network plays a role for labour market outcomes. If other members of an individual's caste are unemployed, they may be less valuable source of information regarding employment opportunities and can exert negative peer pressure for job searching. On the other hand, if other members of an individual's caste are employed, they are likely to provide useful information about job openings and serve as role models. Understanding the effect of social network on employment is important for the policy designed to eliminate the persistent social inequality in India. We use data from a national representative survey and construct district level caste based social network measures in order to study the network effect on employment probability.

Rest of the paper is organised as follows. Section 2 introduces the methodology used to measure social networks and their impact on employment. Section 3 describes the data. Estimation results are presented in Section 4. The last section concludes the paper.

## 2. Methodology

Suppose that the true model governing the probability that an individual is employed is given by

$$\Pr(empl_{ijk} = 1) = netw_{ijk}\alpha^* + X_i^*\beta^* + Y_j^*\gamma^* + Z_k^*\delta^* + u_{ijk} \qquad [1]$$

where i indexes individuals, j indexes districts, k indexes castes, $empl$ is a dummy variable indicating employment, $netw$ measures the information and social pressure from contacts, $X^*$ are observed and unobserved personal characteristics, $Y^*$ are observed and unobserved district characteristics, $Z^*$ are observed and unobserved caste characteristics, and $u$ is an error term.

Measuring *netw* raises difficulties. Few data sets contain information on actual contacts. Moreover, individuals choose their contacts, exacerbating omitted variable biases. For example, an individual with many employed friends may be different from one who has few employed friends. Thus, estimation of this model poses two potentially interacting problems: measurement and omitted variable biases.

Often mean neighbourhood characteristics are used to proxy for networks. This implicitly assumes that contacts are randomly distributed within neighbourhoods. However, this approach suffers from what Manski (1993) calls the “reflection problem.” Does individual behaviour depend on the behaviour or characteristics of the group (social effects), or do individuals in a group behave similarly because they are subject to the same shocks (correlated effects)? The reflection problem can be viewed as a manifestation of two related omitted variables biases. (1) Omitted personal characteristics may be correlated with mean neighbourhood characteristics. For example, individuals living in bad areas may be less ambitious. (2) Omitted group characteristics may be correlated with mean neighbourhood characteristics. For example, an area with prosperous economy may increase an individual’s probability of being employed as well as the employment rate in the area. Both these biases are likely to be positive, resulting in an overestimate of the network effects.

A variety of social network measures is used in empirical studies of network effect. International evidences suggest that social networks tend to sort individuals along dimensions such as ethnicity, religion, language, age and locality (Alba, 1990; Bakalian, 1993; Lin, 2001; Waldinger, 1996). Ethnic group is often an important source of connections for members of the group who are entering the labour market for the first time or who are changing jobs. Locality is also important dimension in social network formation since individuals are more likely to spend time with individuals living in the same locality.

We use caste within a district as our measure of social network. By the constitutional mandate castes in India are grouped in four categories for the purpose of job reservation and political representation policies. They are Upper caste, Scheduled tribe, Scheduled caste and Other backward caste. This identification is also used in economic literature to study caste based discrimination (see Banerjee at al. 2009 and Ito 2009). Though these broad categories include a diverse list of castes in them, within a small geographical location they tend to be homogenous in terms of language or dialect, religion and culture. We use district level network measures in order to exploit this similarity. More precisely, all individuals belonging to a scheduled tribe in district X are considered to be in same network. This network is distinct from the network for individuals belonging to a scheduled caste in the same district. It is important to note that this network measure stands for potential contacts, not actual contacts, since we do not have information about peer networks. We use district fixed effect as well as caste fixed effect in order to minimize omitted variable biases at these two levels.

We define the network measure as:

$$netw_{ijk} = C_{jk} * \overline{empl}_{jk}$$

where $C_{jk}$ is the density of caste $k$ in district $j$ and $\overline{empl}_{jk}$ is the mean incidence of employed individuals in caste $k$ in district $j$. Notice that the network measure incorporates two dimensions of a network - $C_{jk}$ provides a measure of potential contacts one has in one’s district (quantity dimension) and $\overline{empl}_{jk}$ measures employment status of one’s network contacts (quality dimension). Ideally, $\overline{empl}_{jk}$ is the mean incidence of employed individuals in caste $k$ in district $j$ excluding individual $i$.[1]

Given the network measure the estimation equation takes the following form:

$$\Pr(empl_{ijk} = 1) = (C_{jk} * \overline{empl}_{jk})\alpha + X_i\beta + g_k + d_j + \varepsilon_{ijk} \qquad [2]$$

where $g_k$ and $d_j$ are caste and district fixed effects, respectively. The interaction term in the equation provides a measure of the influence of social network on an individual's probability of being employed. A positive estimate of $\alpha$ would imply that social network effect is at work. Given that the dependent variable is binary, we use a probit model for estimation.

Age is often considered to be an important factor in network formation since people tend to spend more time with others of roughly same age. Loury (2006) and Montgomery (1991) maintain that for labour market information the contacts that matter most are the individuals who are slightly older than the potential employee. Burns et al. (2009) use age-language cohort in their network measure. In our estimation we use another network measure that account for caste-age cohort. In our caste-age based network measure we use five age cohorts - 15-24, 25-34, 35-44, 45-54, and 55-65 years. Defining $k$ as a caste-age cohort, the quantity dimension of this measure reflects the density of group $k$ in district $j$. And the quality dimension reflects the mean incidence of employed people in the caste-age cohort $k$ in a district. We also estimate the model in [2] separately for male and female in order to study whether there is difference in network effects in theses two groups.

The caste based network variables used in this study are constructed at district level. They correspond to 585 districts and 4 castes in India. The argument and evidence in literature of the importance of ethnic identity and geographical vicinity in the formation of networks underlie the construction of caste based network at district level. Arguably, even district is a broad geographical category in India. In order to check robustness of our results we also use urban block or rural hamlet level measures of caste based network (more in Estimation Results section).

## 3. Data

We use data from the 61st round of the Employment and Unemployment Survey in India carried out by the National Sample Survey Organization during the period July 2004 – June 2005. Data are collected on employment situation and socio-demographics of individuals from a nationally representative sample of households spanning all the 28 states and 7 union territories. Since we are interested in economically active population, we limit the sample to individuals aged 15-65 years and exclude non-participants in labour force.

Table 1 presents descriptive statistics for the variables used in our analysis. The sample consists for 407,071 individuals from 585 districts. There are 14% individuals from scheduled tribe, 16% from scheduled caste, 38% from other backward caste and 32% from upper caste. The first column of the table shows that 54% of the sample is employed. Average age in the sample is 36 years. There are 49% female and 77% married individuals. There are 35% individuals without any formal education, 23% with primary education, 27% with secondary education and 14% with higher secondary or higher education. An average household in the

[1] However, this measure may reflect unobserved characteristics that an individual has in common with people from the same caste living in the same district, introducing an omitted variable bias. In order to avoid this, in a robustness check we use state or union territory level mean incidence of employed people in a caste. We find qualitatively similar results to those obtained using district level mean incidence of employed individuals in a caste.

sample consist of 6 household members. Around half of the adult members of a household are employed. The sample consists of 68% individuals from rural areas.

Last four columns of Table 1 give break down of these characteristics for four castes. We find that scheduled caste have higher employment rate than other castes. Higher female labour force participation in this caste might be behind this difference. The fact that upper caste have the lowest employment might appear counter intuitive given the disadvantage hypothesis of non-upper castes vis-à-vis upper caste. In this study we do not make distinction between employments in different sectors. A vast majority of non-upper caste individuals are employed in stagnant agricultural sector or in informal sector where disguised employment or underemployment is rampant. Rather we focus on the effect of social network on employment, which has obvious implication for sectoral segregation that deserves further investigation.

In terms of other characteristics – age, female, married and household size – the four castes are very similar. As mentioned, the percentage of individuals living in rural areas for lower castes is much higher than the percentage for upper caste. And, a striking difference among castes is apparent in the distribution of education. The percentage of individuals with no formal education is much higher in scheduled tribe, schedule caste and other backward caste than in upper caste. Similarly, the percentage of people with more than higher secondary education is much higher in upper caste than in other castes.

## 4. Estimation Results

We estimate a probit model for employment. As in equation [2], the right hand side variables include fixed effects for castes and fixed effects for districts, demographic controls, and a measure of network (interaction between contact availability and employment rate of reference group). In addition, we also include contact availability among explanatory variables. The estimation results for the caste and caste-age based networks for the full sample are presented in Table 2. The demographic controls include age, age-squared, sex, marital status, four dummies for education, household size, fraction of adult household members employed, and a dummy for rural area. Since the network variable is constructed from individual data, its inclusion in individual regression can potentially bias standard errors because of correlation among error terms across individuals in a group. In all estimation presented below, the standard errors are corrected for this clustering in network variable.

In Table 2, the estimation titled "Caste based network" uses contact availability and network on the basis of caste and employment rate in a district, whereas "Caste-age based network" uses the measures constructed on the basis of caste and employment rate in a caste-age cohort in a district. The age cohorts used are – 15-24, 25-34, 35-44, 45-54, and 55-65 years.

Now going through the estimates for demographics, we find that estimated coefficients have expected sign. Higher age increases employment probability. The negative and significant estimate for age-squared term shows that the effect of age on employment probability is nonlinear. The variable for marital status is not significant. This variable is likely to be endogenous if network increases the probability of being married. Nevertheless, the inclusion of this variable is more likely to serve as control for unobserved characteristics. This implies that the network effects we find after inclusion of this variable are likely to be underestimates of the true effects. Therefore finding evidence of networks in spite of controlling for this variable only strengthens the case for the importance of network. The effect of education is not monotonic. Medium level of education is associated with unemployment as compared to the omitted group of no formal education. However, there is indication that education increases employment probability after higher secondary level of education though the coefficient appears to be insignificant. Female are less likely to be employed than male. Higher the fraction of employed adults in an individual's household higher is his/her own employment probability. This apparently reflects information spill-over effect of close-knit ties.

Turning to the network estimates, we find that network has a strong positive effect on the employment probability. The sign of the estimate is positive in both specifications. The negative and significant effect of contract availability warrants a few comments. Apparently, it seems to be driven by crowding out effect – higher the number of people of same caste lower is the likelihood of being employed. However, the contact availability variable is also present in the network variable. Hence its effect on employment will depend on the value of the other variable in the interaction term. Simple arithmetic using coefficients from the first estimation reveals that the effect of contact availability, (-1.954+3.712*employment rate of the caste in a district), is negative when the employment rate is less than 52%. For employment rate higher that 52%, the effect of contact availability is positive. Using the second estimates we find a switching point at 53% of employment rate.

The coefficient of network variable is not easy to interpret since the variable is constructed by multiplying two ratios. The effect of employment rate of a caste would depend on the value of contact availability of an individual of that caste. Given contact availability, the estimate indicates that higher the percentage of contracts employed higher is the individual employment probability. To provide an intuitive meaning of the magnitude of the network effect we calculate the change in predicted probability of employment following a policy shock. We consider a policy shock that increases the employment rate over the economy by one percentage point and ask how much network effects would magnify a policy shock affecting employment. Using estimates from caste-based network we find that network magnifies the effect of such policy shock by around another half percentage point. The caste-age based network estimates indicate that the magnifying effect is more than one percentage point. Using language-age based network measures for South Africa, Burns et al. (2009) also find such magnifying effect of a policy shock for different language groups.

Arguably district is a broad category to capture social network in India. We check the robustness of our results using further disaggregated measure of networks. The survey collects data from two or more urban blocks and hamlets from rural areas in a district. We construct network measures at urban block/hamlet level. Our results remain similar to the district level estimates, though network effect appears stronger at more disaggregated level. However, we are not able to use urban block/hamlet level fixed effects in this robustness check as that would imply incorporating more than one thousand dummy variables in a non-linear model. We rather run the model using district level fixed effects as before.

In Table 3, we present estimates of the district level networks for male and female. Some interesting differences are observed for these two groups. Network effect is much stronger for male than for female. It might indicate that relevant network for female is much localised than for male. Another important difference between these two groups is that married men are more likely to be employed and married women are less likely to be employed. Effect of education shows different patterns for male and female. The coefficient of more than higher secondary education is negative for male and is positive for female. This may reflect the difference in continuation in education and search for better job among young male and female. This supposition is partially supported by the fact that the effect of age on employment is higher for female than for male among young individuals. In fact, if we limit the male sample to more than 24 years olds the higher education category is not significant anymore. It also interesting to note that female are more likely to be employed in rural areas whereas male are less likely to be employed in rural areas.

In our robustness check using hamlet level network measures we find further interesting difference between male and female. For male individuals, the magnitude of network coefficient slightly decreases as we move from district level estimates to hamlet level estimates. However, the coefficient increases in magnitude for female as we move from district level to hamlet level. This might be indicative of the fact that the relevant network for female is more localised than for male. However, at both levels the network effect is always stronger for male than for female.

Summing up our results, we consistently find that social network is an important determinant of employment in India. This finding is robust for both district level and further local level network measures. The network effect is present for employment of both men and women, though with different force for these groups. It is important to mention that an individual's social life consists of participation in different over-lapping networks. And the participation itself may be function of an individual's socio-economic status. Our data, as most survey data, do not allow us to identify the different networks an individual belong to and also the level of reach in a given network. Rather, assuming that individuals in same caste and geographical location are more likely to interact and exchange information, we construct network measures along these two dimensions – caste and location. As information about actual contacts in large scale survey are unlikely to come out in near future, we believe that our findings shed light in an area that has received little attention in economic literature concerning Indian labour market.

## 5. Conclusion

In this paper we have estimated the effect of social network on employment in India. Using caste based social network measures and controlling for caste and district fixed effects we find that networks exert a significant effect on employment probability. The results indicate that higher the number of employed people in an individual's caste higher is his/her own chance of being employed. As our results indicate, informational externalities provided through other employed members of household are important determinant of employment. Finding a significant network effect after controlling for this factor underlines the importance of 'weak-ties' for labour market prospects. We find that social network has the potential to magnify policy effect. The results have important implication for policies designed for positive labour market discrimination for disadvantaged castes. The results imply that in addition to direct effect of policy on employment, the network effect generate further positive externality. The other important implication of our finding is that the policy should focus on local level in order to generate additional benefit through network channels. In fact the job reservation policy for men and women should be at different levels considering that the relevant network for women may be more localised than for men.

In this paper, we have studied the effect of social network on employment. However, much of the social inequality in India is manifested in the lower representation of non-upper castes in good jobs or in growing sectors and in the occupational segregation along caste lines. An avenue for future research would be to investigate to what extend the persistence of these phenomena is explained by caste based social networks.

## References

Akerlof, G. A. (1980). A theory of social custom of which unemployment may be one consequence, *Quarterly Journal of Economics* 95, pp. 749-775.

Alba, R. D. (1990). *Ethnic identity: the transformation of white America*, New Haven CT, Yale University Press.

Bakalian, A. (1993). *America-Americans: from being to feeling American*, New Burnswick, NJ, Transaction Publishers.

Banerjee, A., Bertrand, M., Datta, S. and Mullainathan, S. (2009). Labor market discrimination in Delhi: evidence from a field experiment, *Journal of Comparative Economics* 37, pp. 14-27.

Bertrand, M., Luttrmer, E. F. P. and Mullainathan, S. (2000). Network effects and welfare cultures, *Quarterly Journal of Economics* 115, pp. 1019-1055.

Burns, J., Godlonton, S. and Keswell, M. (2009). Social networks, employment and worker discouragement: evidence from South Africa, *Labour Economics* doi:10.1016/j.labeco.2009.08.007

Coleman, J. S., et al. (1966). *Equality of educational opportunity*, Washington DC, Government Printing Office.

Ito, T. (2009). Caste discrimination and transaction cost in the labor market: evidence from rural north India, *Journal of Development Economics* 88, pp. 292-300.

Lin, N. (2001). *Social capital: a theory of social structure and action*, Cambridge and New York, Cambridge University Press.

Loury, M. (2006). Some job contracts are more equal than others: earnings and job information networks, *Journal of Labor Economics* 24, pp. 299-318.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem, *Review of Economic Studies* 60, pp. 531-542.

Montgomery, J. D. (1991). Social networks and labour-market outcomes: towards an economic analysis, *American Economic Review* 81, pp. 1408-1418.

Waldinger, R. D. (1996). *Still the promised city?* Cambridge MA, Harvard University Press.

## Table 1: Descriptive statistics of the sample

| | All | Scheduled tribe | Scheduled caste | Other backward caste | Upper caste |
|---|---|---|---|---|---|
| *mean values* | | | | | |
| Employed | 0.54 | 0.55 | 0.61 | 0.53 | 0.51 |
| Age | 35.63 | 35.23 | 34.85 | 35.54 | 36.30 |
| Female | 0.49 | 0.51 | 0.49 | 0.50 | 0.49 |
| Married | 0.77 | 0.74 | 0.77 | 0.78 | 0.77 |
| No formal education | 0.35 | 0.40 | 0.48 | 0.38 | 0.25 |
| Primary education | 0.23 | 0.27 | 0.24 | 0.24 | 0.22 |
| Secondary education | 0.27 | 0.25 | 0.21 | 0.27 | 0.31 |
| Higher secondary education | 0.06 | 0.04 | 0.03 | 0.05 | 0.08 |
| More than higher secondary education | 0.08 | 0.05 | 0.04 | 0.06 | 0.14 |
| Household size | 5.84 | 5.67 | 5.62 | 5.97 | 5.88 |
| Fraction of adult household members employed | 0.52 | 0.53 | 0.59 | 0.52 | 0.49 |
| Rural | 0.68 | 0.81 | 0.72 | 0.69 | 0.60 |
| Number of observations | 407,071 | 55,793 | 66,945 | 153,549 | 130,784 |
| Percentage of sample | 100.00 | 13.71 | 16.45 | 37.72 | 32.13 |

## Table 2: Estimates of the probit model for employment – caste and caste-age networks

| | Caste based network | | Caste-age based network | |
|---|---|---|---|---|
| | Estimate | S. E | Estimate | S. E. |
| Contact availability | -1.954 *** | 0.103 | -1.613 *** | 0.036 |
| Network (interaction b/w contact availability and employment rate of contacts) | 3.712 *** | 0.195 | 3.095 *** | 0.056 |
| Age | 0.114 *** | 0.002 | 0.078 *** | 0.002 |
| Age squared | -0.001 *** | 0.000 | -0.001 *** | 0.000 |
| Female | -1.928 *** | 0.017 | -1.936 *** | 0.013 |
| Married | 0.007 | 0.014 | -0.011 | 0.012 |
| Primary education | -0.089 *** | 0.011 | -0.100 *** | 0.009 |
| Secondary education | -0.241 *** | 0.012 | -0.242 *** | 0.010 |
| Higher secondary education | -0.248 *** | 0.020 | -0.240 *** | 0.016 |
| More than higher secondary education | 0.011 | 0.017 | 0.017 | 0.015 |
| Household size | -0.074 *** | 0.002 | -0.075 *** | 0.001 |
| Fraction of adult household members employed | 0.574 *** | 0.022 | 0.569 *** | 0.014 |
| Rural | -0.023 ** | 0.010 | -0.022 *** | 0.008 |
| Caste fixed effects (3 dummies) | | Yes | | Yes |
| District fixed effects (584 dummies) | | Yes | | Yes |
| Log likelihood | | -185280.890 | | -182528.410 |
| Pseudo R-squared | | 0.337 | | 0.346 |
| Number of observations | | 407,071 | | 407,071 |

*Note: *** and ** stand for significance at 1% and 5%, respectively.*
*Standard errors account for the clustering in network measure.*

Table 3: Estimates of the probit model for employment – male and female

| | Caste based network: Male | | | Caste based network: Female | | |
|---|---|---|---|---|---|---|
| | Estimate | | S. E | Estimate | | S. E. |
| Contact availability | -4.426 | *** | 0.166 | -0.833 | *** | 0.041 |
| Network (interaction b/w contact availability and employment rate of contacts) | 5.636 | *** | 0.208 | 2.858 | *** | 0.130 |
| Age | 0.115 | *** | 0.003 | 0.155 | *** | 0.003 |
| Age squared | -0.001 | *** | 0.000 | -0.002 | *** | 0.000 |
| Married | 0.536 | *** | 0.019 | -0.509 | *** | 0.021 |
| Primary education | -0.008 | | 0.015 | -0.191 | *** | 0.014 |
| Secondary education | -0.189 | *** | 0.015 | -0.296 | *** | 0.016 |
| Higher secondary education | -0.307 | *** | 0.022 | -0.118 | *** | 0.029 |
| More than higher secondary education | -0.212 | *** | 0.020 | 0.394 | *** | 0.026 |
| Household size | -0.112 | *** | 0.002 | -0.064 | *** | 0.002 |
| Fraction of adult household members employed | 0.490 | *** | 0.021 | 1.983 | *** | 0.023 |
| Rural | -0.278 | *** | 0.013 | 0.061 | *** | 0.011 |
| Caste fixed effects (3 dummies) | | | Yes | | | Yes |
| District fixed effects (584 dummies) | | | Yes | | | Yes |
| Log likelihood | | | -75591.505 | | | -72982.059 |
| Pseudo R-squared | | | 0.253 | | | 0.369 |
| Number of observations | | | 205,986 | | | 201,085 |

*Note: *** stands for significance at 1%.*
*Standard errors account for the clustering in network measure.*